Responsible Use of Research Content in AI: Importance and Benefits

IEEE
Advancing Technology for Humanity

CONFERENCE PROGRAM: NAVIGATING AI IN SCHOLARLY COMMUNICATION, LIBRARIES BUILD TRUST AND ACCESS
CONCERT / NATIONAL CHENGCHI UNIVERSITY, NOVEMBER 10-11, 2025

Ira Tan, Regional Manager, APAC, IEEE
10 Nov 2025

Introduction

IEEE Perspective

IEEE is the world's largest technical professional organization, dedicated to advancing technology for the good of humanity

IEEE perspective as an organization serving the community

- Embrace of new technologies is **core to our mission**.
- We strive to **serve the needs** of the scholarly research community.
- **Developments in AI** are already impacting scholarly communications.
- **IEEE's role** includes helping guide AI adoption and use in scholarly communications in ways that will *optimize the benefits* and *mitigate the risks* to authors and researchers in our community.

Research Content and AI in Context

The Landscape

Overview of AI applications in scholarly communications

AI applications offer benefits across the scholarly communications cycle

Idea Generation and Literature Discovery

- Smart literature search
- Trend detection
- Knowledge graphs

Experiment Design and Data Collection

- Simulation and modeling
- Automated measurement and lab control
- Data augmentation

Data Analysis and Interpretation

- Pattern recognition
- Statistical rigor
- Visualization

Manuscript Preparation

- Drafting support
- Language refinement
- Equation and code formatting

Peer Review and Revision

- Pre-submission checks
- Reviewer response drafting
- Bias checks

Publication and Dissemination

- Target journal selection
- Metadata optimization
- Multi-format output

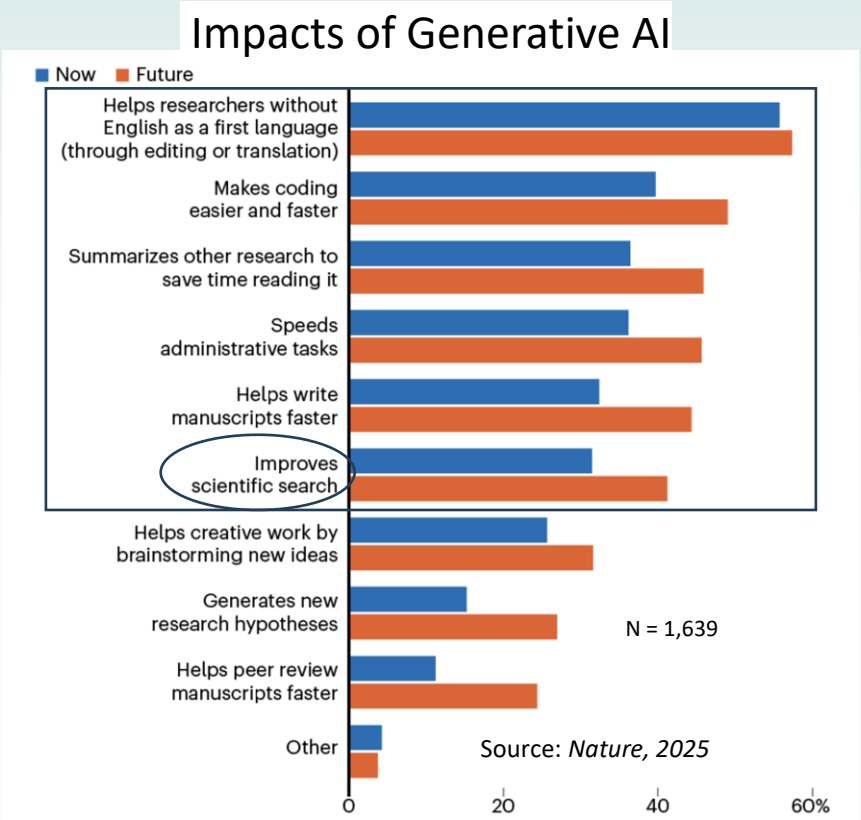
Researchers' perceptions of the benefits of generative AI

Researchers see the most benefits in several aspects of the research process

- **What do you think are currently the biggest benefits of generative AI for research?**
- **In the future, where do you think generative AI will have the biggest beneficial impacts for research?**

Respondents to a *Nature* survey most often highlighted **translation** as a benefit to research *(current and future*), followed by **coding**, **summaries**, **administrative tasks**, **faster writing**, and ***improved search.***

Of less perceived interest (at this early stage) were help with **brainstorming**, **generating hypotheses**, and **faster peer review**.

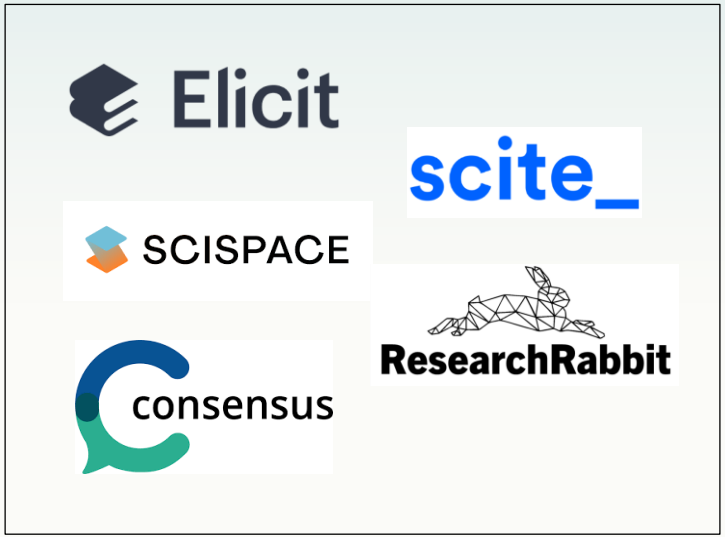


AI tools for scholarly research discovery

The landscape of AI tools that expedite research/discovery is diverse and growing

Established Academic Sites with AI Features

Emerging AI Research Tools



Generative AI Native Apps

Publisher's own vs "native" generative AI platforms

Content used in generative AI outside of publishers' own platforms brings challenges

- Publisher platforms are leveraging LLM technology to create proprietary AI tools and services designed for **specific research applications**, e.g.:
  - Elsevier Scopus AI
  - Clarivate Web of Science Research Assistant
  - Dimensions Research GPT
  - Springer Nature's multiple AI tools
  - IEEE *Xplore* AI Research Suite
- Publisher platforms draw upon their content resources (metadata, structure, curation) to help maintain key values of scholarly communication such as **reliability**, **verifiability**, **attribution,** and **accuracy**.
- However, STM/scholarly content will also continue to be used within the major **general-purpose generative AI platforms** (ChatGPT, Claude, Gemini, Grok etc.) in which *the elements needed for responsible use are largely absent*.

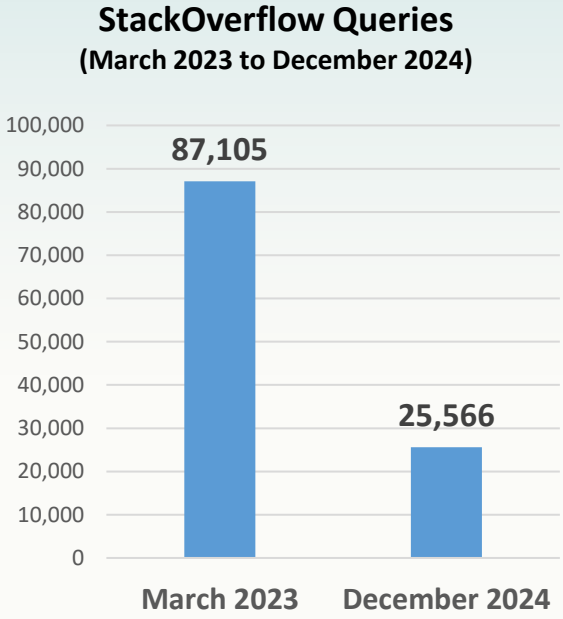
Integration of Research Within AI Platforms

Trends and Challenges

AI use in scholarly research discovery

Generative AI platforms likely will be a large element of STM discovery workflows

- Students and faculty in Science and Engineering are *least likely* to begin their literature discovery process with library resources.[1]
- A 2025 survey found that 21% of higher ed students began their research process by using ChatGPT (18%) or another generative AI tool (3%).[2]
- STM publishers are observing bot traffic from AI platforms outpacing genuine researcher visits.[3]
- StackOverflow (leading Q&A platform for software developers) saw usage plunge by over two-thirds from early 2023 to late 2024, ostensibly as a result of AI chatbots providing answers directly from Stack Overflow.[3]

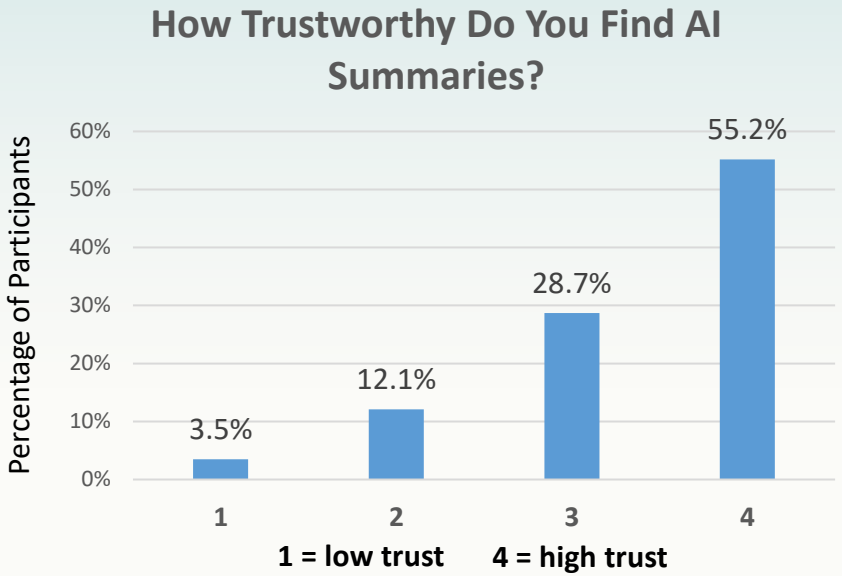


Sources: 1,2 - Sage Technology (2021, 2025); 3 - Scholarly Kitchen

The impact of AI summarization on original content sources

Google's AI Overviews materially decrease traffic to sources of search results

- A new study using a general sample of desktop users found that:
  - Outbound clicks to the underlying content sources **fell by two-thirds** when an AI Overview (AIO) appeared.
  - **7 in 10 searchers** never read past the first third of an AIO.
  - Over half of the participants expressed **"high trust"** of AIO summaries.



Although not specific to scholarly content, these findings indicate a general behavioral challenge to providers of content included in Google search results.

Source: Growth Memo

Problematic issues of AI integration of research content

General-purpose generative AI models could disrupt scholarly communications

Risks	Implications
▸ Declining direct usage of original content sources	• Bypass of library services and resources • Reduced recognition of authorship • Risks to content provider sustainability
▸ LLM inaccuracies and “hallucinations” ▸ Misinformed summarization of original research leading to poor science	• Dilution of trust in scientific knowledge • Potential harm in critical applications
▸ Incomplete/improper attribution and prioritization of peer-reviewed content	• Motivational and career impacts on researchers • Weakening of the scientific enterprise

Toward Responsible AI/Content Solutions

Principles, engagement, advocacy

Principles for incorporating research content in AI tools

Researchers rely on key content values to generate trustworthy insights

Key Principles	Importance
▸ **Verifiability** through proper attribution and citation	• Supports author impact • Ensures trustworthiness and provenance
▸ **Differentiation** between peer-reviewed and non-peer-reviewed content	• Prioritizes validated content • Keeps value of (e.g.) preprints in context
▸ **Inclusion** of rebuttals, corrections, and retractions in content processing and display	• Supports the reliability of scholarly content • Provides for debate and accountability
▸ **Prioritization** of the original publication in references, citations, inference and retrieval	• Ensures display of up-to-date knowledge • Avoids perpetuating errors, misinformation

Source: STM

Principles for incorporating research content in AI tools (2)

Assurance of key rights and protections is also key for the research community

Key Principles	Importance
▸ **Protection of user privacy** through, e.g., transparent prompt management	• Conforms to ethical requirements • Avoids potential privacy-related liabilities
▸ **Respect for copyright and licensing** to ensure adherence to permissions and terms	• Ensures sustainability of scholarly enterprise • Respects rights of authors and stakeholders
▸ **Bias mitigation** and promotion of fairness and impartiality in content selection and processing	• Supports trust in scholarly content • Discourages bad-faith manipulations
▸ **Transparency** for end users and content authors, publishers and owners regarding these rights and protections	• Fosters a constructive scholarly community • Preserves key scholarly content values

Source: STM

Risks of not adhering to key principles of scholarly content

Their absence would expose the advancement of science to significant risks

- Spread of misinformation
- Opportunities for deliberate misinformation
- Erosion of trust
- Safety issues in critical applications
- Undermining of researcher motivation
- Hindering of reproducibility and validation of scientific work

Source: STM

Created with ChatGPT

Approaches to help ensure research content integrity in AI

STM publishers can leverage existing and emerging ecosystems and architectures

Retrieval-Augmented Generation (RAG)

A framework that uses an external system to retrieve relevant information before an LLM gives a response.

- Used by Elsevier, Clarivate, Digital Science, Wiley, IEEE
- Enhances generation with vetted sources
- Provenance-rich retrieval (e.g. via Crossref/OpenAlex metadata)
- Reduces “hallucinations” and improves coverage of niche literature

The Model Context Protocol (MCP)

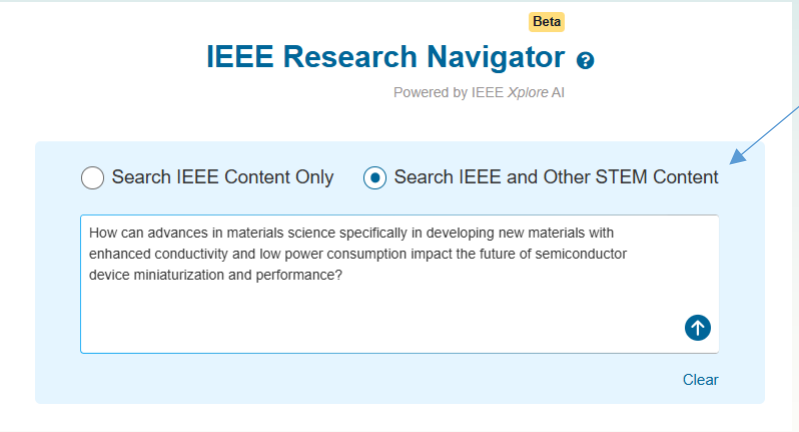
An open standard and framework for communication between LLMs and external tools, data sources, and applications.

- First introduced by Anthropic
- Extends and complements RAG
- Provides a framework for interacting with external tools and managing agents
- Used by Wiley in work with Anthropic to establish best practices for citations and copyright

While RAG focuses on retrieval before *generation, MCP provides a standardized interface for LLMs to request additional information or perform actions* during *the generation process.*

Retrieval-Augmented Generation (RAG) enhances the LLM

IEEE *Xplore* AI Research Suite Example



User prompt searches against a LLM trained on IEEE full text and an expertly curated set of STEM-related metadata and abstracts from OpenAlex

Using RAG, AI overview is returned with inline references for further exploration

IEEE AI Overview

Advances in materials science, particularly in developing new materials with enhanced conductivity and low power consumption, are crucial for the future of semiconductor device miniaturization and performance. The integration of novel materials has been pivotal in advancing Si-based devices, enabling the industry to uphold Moore's Law for both More Moore and More than Moore devices [4]. New materials like two-dimensional materials, transition metal dichalcogenides, carbon nanotubes, and metal oxides hold promise for further scaling endeavors, allowing for the fabrication of high-performance devices without relying solely on Si [4].

Materials such as graphene, carbon nanotubes, and metal oxides are revolutionizing the field of electronics by enabling the development of high-performance devices with exceptional electrical conductivity, thermal management capabilities, and potential for device miniaturization [5]. Graphene, for instance, exhibits… Show More

References:

1. Performance Evaluation Of Emerging Semiconductor Devices For Low Power Applications: A Review
2. Comprehensive Analysis of Advancements and Challenges in Smartphone Semiconductor Technology
3. The Role of Semiconductor Thin Films in Advancing MEMS Sensor Technology
4. Future Materials for Beyond Si Integrated Circuits: A Perspective
5. Nanomaterials in electronics: Advancements and challenges in high-performance devices
6. Revolutionizing Electronics and Technological Interventions in Semiconductor Devices

Takeaways for academic authors and researchers

Generative AI will bring both benefits and challenges to scholarly communications

- Generative AI will provide powerful benefits to authors and researchers across the research life cycle.
- Established publishers and new entrants are harnessing AI capabilities to enable improvements in the efficiency, insight, and impact of scholarly research.
- Powerful and rapidly advancing general-purpose LLM platforms will bring additional, perhaps unimagined benefits to the scientific enterprise.
- *It will be critical for the scholarly communications community as a whole (authors, researchers, publishers, funders, research institutions) to ensure that the leading edge of generative AI capabilities incorporate and maintain the key elements of responsible AI use of research content that undergird research integrity and scientific advancement.*

Strategies for successful AI / research content integration

Librarians and publishers can join to advocate effective integration by AI providers

- **Librarians**
  - In AI vendor selection and negotiation, advocate for the importance of:
    - Content attribution and citation, prioritizing the article of record
    - Differentiation of peer-reviewed and non-peer-reviewed content
    - Transparency of LLM procedures and policies regarding research content
    - Respect for user privacy and licensing of copyrighted content
  - Facilitate dialogue on these issues among authors, users, publishers, and AI providers
- **Publishers**
  - Emphasize how effective research content integration bolsters perceptions of AI platform integrity
  - Ensure availability and deployment of ecosystem resources to signal quality and trusted content
- **Both Together**
  - Educate the scholarly research community (donors, educational institutions, corporate R&D, media) on the importance of responsible use of research content in LLMs and Generative AI.
  - Collaborate on new features/approaches that can harness the power of AI to create new knowledge.

Thank you!